Unemployment Insurance
Tax Data Validation

Web-Based Data Validation Software

Software Tutorial

June 2005

Unemployment Insurance Tax Data Validation Web-Based Software System

TUTORIAL

TABLE OF CONTENTS *PAGE*

TABLE OF CONTENTS (CONTINUED) *PAGE*

1. Introduction and Overview: The Data Validation Program

1.1 Background

State Unemployment Insurance (UI) operations are required to file a series of standardized reports with the Employment and Training Administration (ETA) of the U.S. Department of Labor (DOL). Tax data validation is exclusively concerned with the ETA 581 Contribution Operations Report, a quarterly report covering UI tax operations. Information in the ETA 581 Report is used for a variety of purposes and provides basic information about state UI tax operations.

Since state programs differ significantly within established parameters and states utilize a variety of accounting and data processing arrangements, the issue of the comparability among state reports has emerged. State reporting requirements are standardized, but states use a variety of reporting arrangements and must interpret reporting requirements within the context of their own laws and accounting conventions.

The UI Data Validation (DV) program was established in an attempt to identify and address discrepancies. The program basically requires that states reproduce their reported numbers in a controlled environment, compare these numbers with actual numbers reported to DOL, and correct significant discrepancies. The DV program also requires that states examine a sample of reported cases to verify that the correct information is being counted. They must also perform a series of “sort” operations to verify that records with appropriate state-specific code values being counted.

During the initial stages of the program, PC-based software was produced to facilitate the validation process at the state level and to generate standardized outputs which document state participation in the DV program. These reports establish the level of errors discovered in the validation process. States are said to have “passed” validation when observed error rates are within established tolerance levels.

This PC-based software is being replaced by Web-based software which runs on DOL Sun computers in state UI offices. This computer has an established interface with DOL for report transmission and can retrieve reported numbers from the state data base of reported numbers.

This tutorial provides basic operating instructions for this Sun-based software. One purpose of the tutorial is to assist users familiar with the prior DV system in transitioning to the new web-based Data Validation system. This tutorial provides users with a step-by-step guide for using the web-based Data Validation application.

1.2 Definitions

Certain terms are used in the validation process which have a specialized meaning within the context of the DV program:

1. *Populations*. Populations are sets of transaction records specific to a given kind of activity. “Tax Population one’, for example, refers to the set of transactions which must be counted to satisfy reporting requirements specific to “active employers”. Populations are specific to a given reporting period (e.g., 1st Quarter, 2005) and do not overlap. That means that a given record appears in only one population.

2. *Sub-population*. Each record in a given population record set is assigned to a specific sub-population by the software in the process of loading the data set. Records in sub-populations yield the actual report counts. Each record counted for reporting purposes must be assigned to a sub-population. If a given record does not satisfy the requirements for assignment to a sub-population, this record cannot be counted and is said to be in error.

3. *Extract Files*. These files must be prepared by each state individually and consist of information *extracted* from state production data bases. They are used as input for the DV software and drive all the other functions. The format of extract files must strictly conform to the specifications (record layouts). The software edits these files as part of the loading operation. Before being loaded to the DV software each record must pass syntax and logic edits. Duplicate records are also rejected.

4. *Report Validation (RV).* This refers to the validation of summary counts provided to DOL. Reported numbers must be reproduced in a controlled environment and shown to match reported counts in order to pass validation.

5. *Data Element Validation (DEV)*. This refers to the investigation of samples of records from a given extract file to establish that the information in individual records is accurate and appropriate for reporting purposes, and conforms to federal reporting requirements. DEV for the tax system also requires analysis of extract files using sort routines.

6. *Record Layouts*. These provide detailed information required for the preparation of extract files. Module Three of state-specific Data Validation Handbooks can be used to map the data elements in the record layouts to elements in individual state systems.

1.3 Basic Methodology: Report Validation (RV)

Report validation consists of verifying that report elements are within established tolerance levels. An extract file is prepared for each population and loaded to the software. The loading process identifies records with syntax errors, records with logic errors (which prohibit them from being assigned to a sub-population), and duplicate records. All of these records are diverted to an error table in the database.

The Sun software retrieves reported numbers from the state data base of reported numbers. This is accomplished with reference to the date information entered by the operator when the extract file is loaded. A report is displayed giving the reported numbers, validation numbers derived from the extract file, and the extent of any discrepancy. If the comparison for individual groups of numbers is within tolerance, the state has passed this aspect of validation. The software can be used to forward DV results to DOL where they are stored in a national data base. State staff should save a copy of summary reports as well as a copy of the extract file until the next validation of the same population to satisfy audit requirements.

1.4 Basic Methodology: Data Element Validation (DEV)

A series of small samples are produced by the software from each extract file. The Tax validation system uses “minimum” samples which consist of two records per sub-population within an extract file. Record included in these samples are The software also provides a sort utility which is used to establish what code values are in the extract file for a given Tax population for certain key variables.

The sample frame for each sample consists of a set of specific sub-populations within a given extract file. Records included in a sample are displayed along with the data elements to be validated in a data entry screen. Investigators review each record, identify any elements found to be erroneous, and data-enter this information into the system. The software calculates an error rate for the entire sample. If the DEV investigation establishes errors within established error tolerance levels, the state has passed. Summary reports of DEV validations can be forwarded to DOL. State staff should save a copy of summary reports as well as a copy of the extract file until the next validation of the same population to satisfy audit requirements. States are also encouraged to save individual sample worksheets as documentation of the DEV process.

NOTE: Validations must be performed for specific reports which have been provided to DOL by states. Thus, extract files must be prepared for time periods which match the time period covered by an actual report (e.g., the ETA 581 Report for the 1st Quarter of 2005).

NOTE: DEV and DV validations, including summary validation of counts and dollar amounts, should be performed using the same extract file.

1.5 Wage Item Validation

Validation of data from the Tax program also requires validation of incoming information provided by employers pertaining to wages paid to individuals on a quarterly basis (Wage Records). This information is not included in extract files. Validation of wage record information requires a review of incoming information and a comparison of reported numbers on the ETA 581 Report with re-constructed counts. This information is key-entered into the software and forwarded to DOL.

1.6 System Navigation

The Sun-based DV software is a web-based application with certain characteristics that the user should be aware of, particularly with reference to the prior, PC-based application.

- *Multiple users*. The Sun-based application does not support multiple, concurrent users.

- *Execution Speed*. The Sun-based software may take longer to execute a given command than the PC-based application.

- *Time Out*. The system will exit the user if it does not observe any activity for thirty minutes. During the extract loading operation, the time-out parameter is set for four hours. This allows very large data sets to be loaded without interruption. The user should perform "save" operations frequently if there is a danger of work being lost due to operator inactivity.

- *Exit from Screens*. The user can exit from a given window within the application through use of the "X" in the upper right corner of the window. Be aware that the "X" at the extreme upper corner of the screen will exit the user from the entire application. This will require the user to sign on again and may result in lost data.

- Use of the *Back* option. The browser used for this web-based application has a *Back* button which allows the user to return to a previous screen. Users should be aware that use of this button may result in unexpected results. For example, when executing a query from the **Tax Selection Criteria** screen for a given population, exiting from the resulting screen, and then submitting the same query for a different population, the software may display the prior screen rather than generating a new one. This problem can be avoided by use of the "X" exit option instead.

- *Next link*. Screens that are unable to display all selected records at once because there are too many to fit on the screen have a link labeled *Next* at the bottom of the screen. This takes the user to the next screen display.

- *<Control End> and <Control Home>*. <Control End> will take you immediately to the bottom of any screen and <Control Home> to the top.

- *Report Quarter and Year*. The date information required on the initial screen is needed only for the import function. Once a population is loaded, the user may choose any other function without supplying these dates.

- *Navigation within a Function*. The user may move around within each part of the software through the use of sequential screens. This is explained in the relevant parts of the document.

- *Print Function*. Certain screens have a print button at the bottom of the screen. The print function is particularly important when investigators are verifying DEV cases since using a print of sampled cases may be more realistic than attempting to investigate records and perform data entry at the same time. If print requests result in truncated screens, try resetting the print utility using the "landscape" option. In certain cases it may be necessary to save the file in Word or some other utility and print from there. System administrators should be able to assist if problems are encountered.

- *Save and Save As Functions*. Use of the *Save* button results in a data set being saved internally in the DV application. Saved files for a given population are overwritten when a new extract file for that population is loaded. The *Save As* button allows the user to save data sets at a location outside the DV software. Data sets saved outside the application are not affected by loading new extract files. The *Save As* button can also be used to save summary reports, extract files, and DEV worksheets to satisfy audit requirements.

- *Help functions*. The application has *Help* links on certain screens. Clicking on this option displays information relevant to the data displayed or processes controlled by this screen to assist users in the validation process.

- *Links on the* **Tax Selection Criteria** *Screen*. Labels on the five data-entry variables on the **Tax Selection Criteria** screen (*Report Quarter, Year, Population, Report Due Date,* and *Choose a function*) serve as links to additional information. Clicking on *Population* also retrieves record layout information for the extract files specific to each tax population.

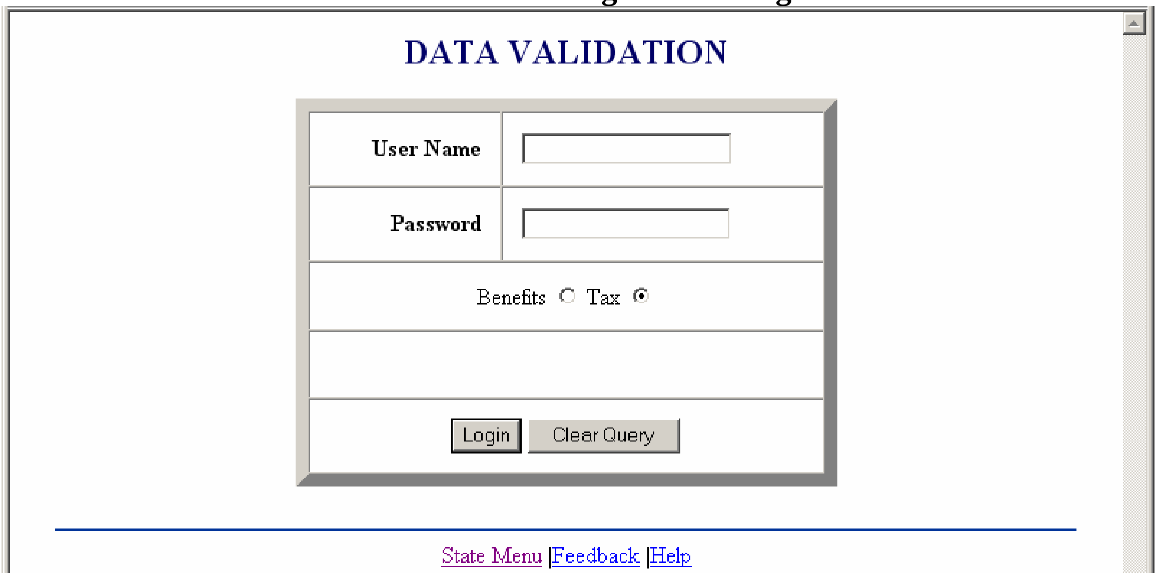
The remainder of this document provides explanations, and sequential steps state staff should follow when using the Sun-based DV software.

2. Logging on

1. Select *Data Validation* from the State Menu**.**

2. Select *Main Login.*

Figure 2.1 - Login Screen



3. Enter the *User Name* Example: Ruth

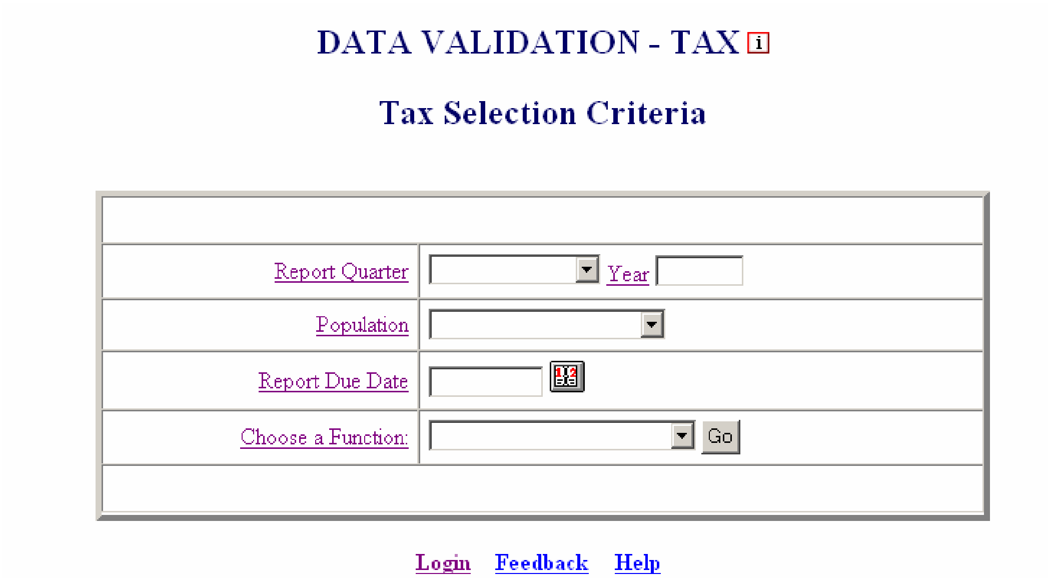
4. Enter the *Password* Example: XNB

5. Select *Tax*

6. Click on the *Login* button

7. **The Tax Selection Criteria Menu** (Figure 2.2) will be displayed. This is the main menu for selecting all validation application functions.

Figure 2.2 – Tax Selection Criteria Screen



8. Select a *Report Quarter* and *Year* from the drop–down box.

9. Select a *Population*. This is used to identify one of the five tax populations.

10. Select a *Report Due Date* from the calendar available as a drop-down box.

NOTE: *Report Due Date* is a state-designated date after which the state assesses penalty or late charges to employers. It is only relevant to Tax Population 2.

NOTE: The user can click on the underlined items displayed on the **Tax Selection Criteria** screen to retrieve information about these variables. Clicking on *Population* will retrieve record layouts for these populations.

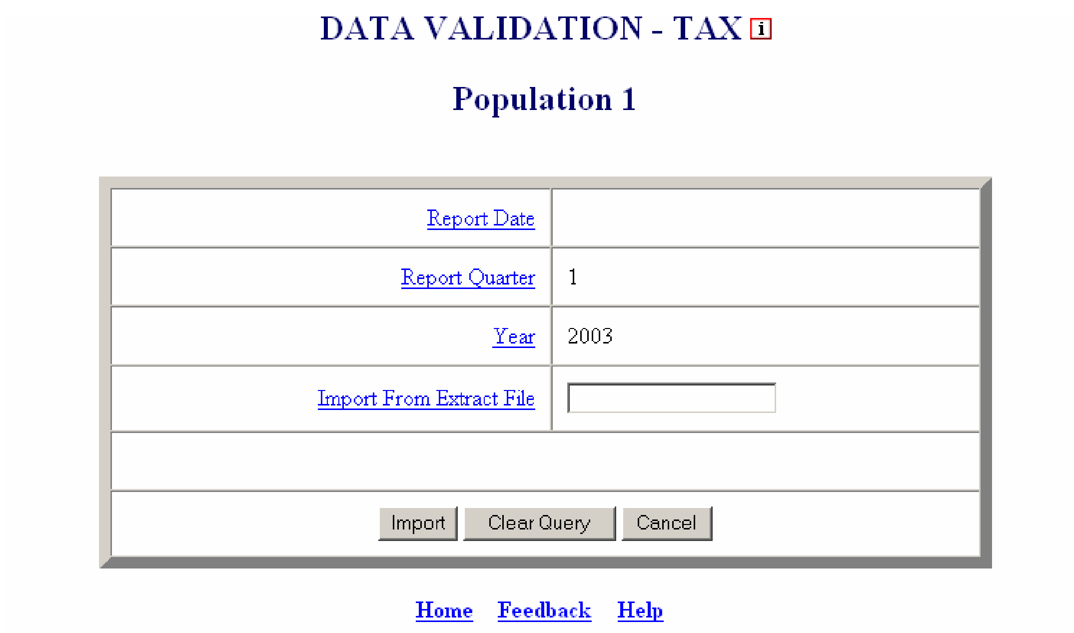
3. Loading Extract Files

Extract files are prepared by state technical staff from state production systems. The format of each of the five extract files used in tax Data Validation are provided in Appendix A of the Data Validation Handbook, and may also be retrieved by clicking on the *Population* link on the **Tax Selection Criteria** screen.

Incoming extract files are subjected to various edits to identify 1) syntax errors, 2) logic errors, and 3) duplicate records. Information is provided in this Section describing how the user can determine which records ended up where.

1. Select *Import Data* from the *Choose A Function* drop-down box (see Figure 2.2) and click *Go***.** This will retrieve the screen illustrated below:

Figure 3.1 – Import Data Screen



2. Enter the full path where the file is located and the name of the extract file into the Import From Extract File field (example: /opt/dv/data/popw.txt).

   NOTE: The location of extract files to be loaded is at the discretion of the system administrator.

3. Click on the *Import* button to load the extract file into the system. This will take you to the **View Import Messages** screen for information on loading processing.

NOTES:

- The software allows multiple populations to be resident in the application at the same time. Each population loading should be completed before loading an additional population. If the same population is loaded a second time, the new data set will over-write the former.

- Re-loading the same population will produce identical results with reference to summary (RV) validation. However, each sample drawn from an extract file tends to generate a unique set of records.

- Load times vary depending on the number of records in the extract file. The time-out parameter is set for four hours to allow ample time for loading large files. Most populations complete loading within a few minutes. Load times are affected by the size of the file, the population being loaded, and the number of error conditions encountered during loading.

- If you wish to cancel a load in progress, use the *cancel* button on the **Import Data** screen. If you return to this function from elsewhere in the application, you must identify the population and provide start and end dates for the date range.

3.1 Import Messages

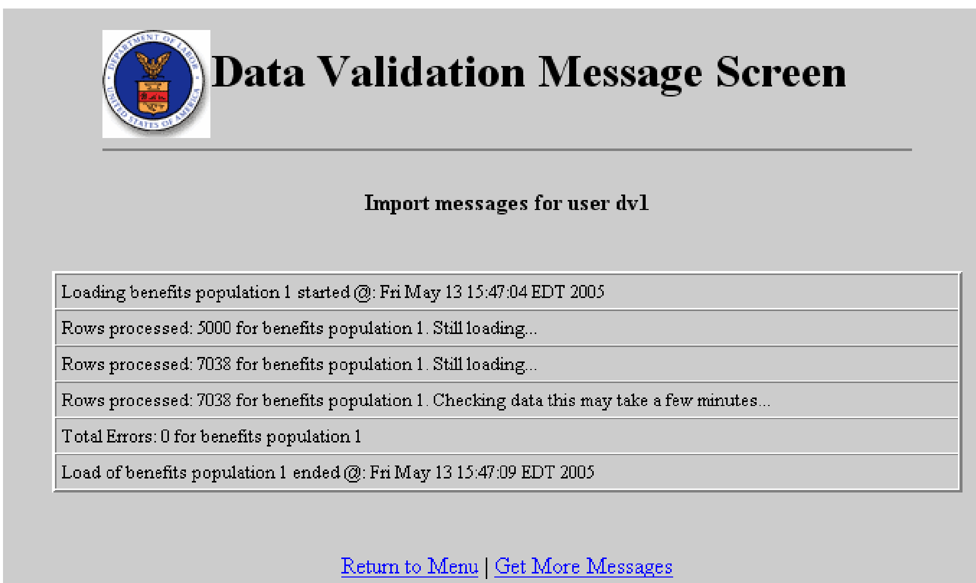
Information about the status of data sets which have been loaded to the system is available at various locations.

1. From the **Tax Selection Criteria** screen select a population that has been loaded or is in the process of loading.

2. Select *View Import Messages* from the *Choose A Function* drop-down box (See Figure 2.2 **Tax Selection Criteria** screen).

This screen displays information including the time the load started, the time the load finished, the number of rows processed, and the number of records with errors detected during the loading routine.

Figure 3.2 – View Import Messages Screen



When the load is complete, you will get a count of the number of rows that loaded and also a count of the number of records with errors encountered in the load process. The number of rows processed provided a count of the number of records in the complete extract file.

3. When extract files are loaded, the software reads each record to ensure that all fields are valid based with reference to specifications provided in the Record Layout (Appendix

A of the Data Validation Handbook, or click on the *Population* link on the **Tax Selection Criteria** screen). Duplicate records and any records with invalid data or missing data in mandatory fields are loaded to an error file rather than to the Source Table. Records in the Source Table are used by the software for all subsequent functions.

NOTE: Messages displayed on the Import Messages Screen pertain to the import of a specific population. These messages are available during the loading operation, but are not available after the file has been loaded and the operator has left this screen. Information about previous population loads is not available. Users have the option of printing this screen for future reference.

3.2 Error Conditions in Incoming Files

Three kinds of error conditions are detected during the import and loading process:

1. *Syntax errors.* This refers to records which are not formatted according to instructions in the population-specific Record Layouts. An alpha character in a numeric field, or an empty field for a required variable, are conditions which are identified as syntax errors.

2. *Parsing errors*. All records in an extract file must be assigned to a sub-population. This parsing operation requires that information in fields within individual records be used to assign records to sub-populations. If sufficient information is missing, or if there are logic errors within the record, the record cannot be assigned to a sub-population and is loaded to the error file. An example of a logic error would be a case where a record providing information about a new claim has the value '0' in the MBA field.

3. *Duplicate records*. Records which are found to be duplicate are not loaded to the Source Table.

The **View Errors** Screen allows the user to view the contents of records found to have syntax or logic errors (which prohibit parsing) during the loading operation. This screen is accessed by using the View Error option from the **Tax Selection Criteria** Screen.

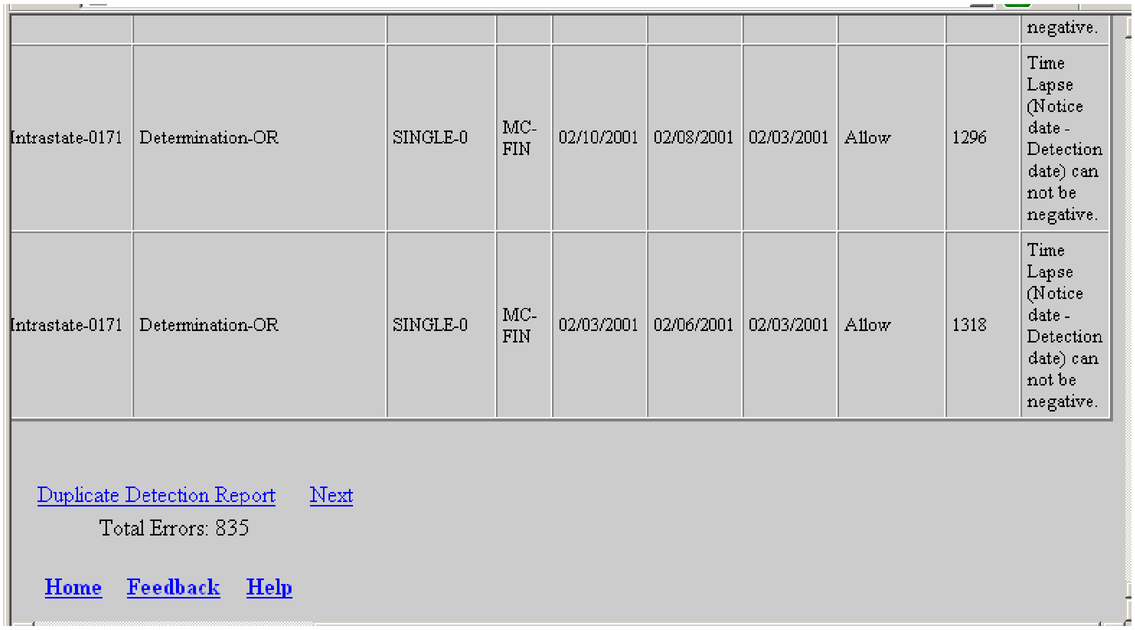
Figure 3.3a - View Errors Screen

Errors for Benefits population 5

Obs	SSN	Issue #	TypeUIProgram	Program Type	Inter/Intrastate	Determination/Redetermination	Single/Multi Claimant	Issue Type	First Week Affected	Detection
3	3 - -	2	Regular UI-UI	UI-MN	Intrastate-0146	Determination-OR	SINGLE-0	VL-QTR	02/10/2001	02/12/2001
85	85 - -	2	Regular UI-UI	UI-MN	Intrastate-0172	Determination-OR	SINGLE-0	VL-MED	01/03/2001	01/10/2001

This figure illustrates the upper portion of the View Errors Screen. Individual data fields from each record classified as being in error are displayed.

Figure 3.3b – View Errors Screen



									negative.
Intrastate-0171	Determination-OR	SINGLE-0	MC-FIN	02/10/2001	02/08/2001	02/03/2001	Allow	1296	Time Lapse (Notice date - Detection date) can not be negative.
Intrastate-0171	Determination-OR	SINGLE-0	MC-FIN	02/03/2001	02/06/2001	02/03/2001	Allow	1318	Time Lapse (Notice date - Detection date) can not be negative.

The bottom portion of the View Errors screen is provided above. Note that the type of error which resulted in the record being rejected is provided in the right column, labeled “Error Message.” The count at the bottom of the screen provides the number of records in the error table for a given population.

3.3 Duplicate Records

The **Duplicate Detection Report** Screen provides an inventory of records found to be duplicates in the process of loading an extract file. This Screen is accessed by clicking on the *Duplicate Detection Report* field at the bottom of the **View Errors** Screen (see above).

Figure 3.4 – Duplicate Detection Report

Duplicate Detection Report for population 5

Obs	SSN	Issue #	TypeUIProgram	Program Type	Inter/Intrastate	Determination/Redetermination	Single/Multi Claimant	Issue Type	First Week Affected	Detection
21	21	1	REGULAR UI-UI	UI-MN	INTRASTATE-0141	DETERMINATION-OR	SINGLE-0	VL-QGC	2000-12-09	2000-12-26
21	21	1	REGULAR UI-UI	UI-MN	INTRASTATE-0477	DETERMINATION-OR	SINGLE-0	MC-FIN	2001-01-03	2001-01-17
33	33	1	REGULAR UI-UI	UI-MN	INTRASTATE-0145	DETERMINATION-OR	SINGLE-0	VL-MED	2001-01-03	2001-01-16
33	33	1	REGULAR UI-UI	UI-MN	INTRASTATE-0170	DETERMINATION-OR	SINGLE-0	MC-FIN	2001-01-06	2001-01-08

The right-most column of this report identifies the condition which resulted in individual records being identified as duplicate, e.g., "SSN and Issue ID."

There is a link at the bottom of this screen to *All Errors*. This displays a table containing all records identified as being erroneous in a given extract file (syntax, parsing and duplicate errors).

3.4 Source Table

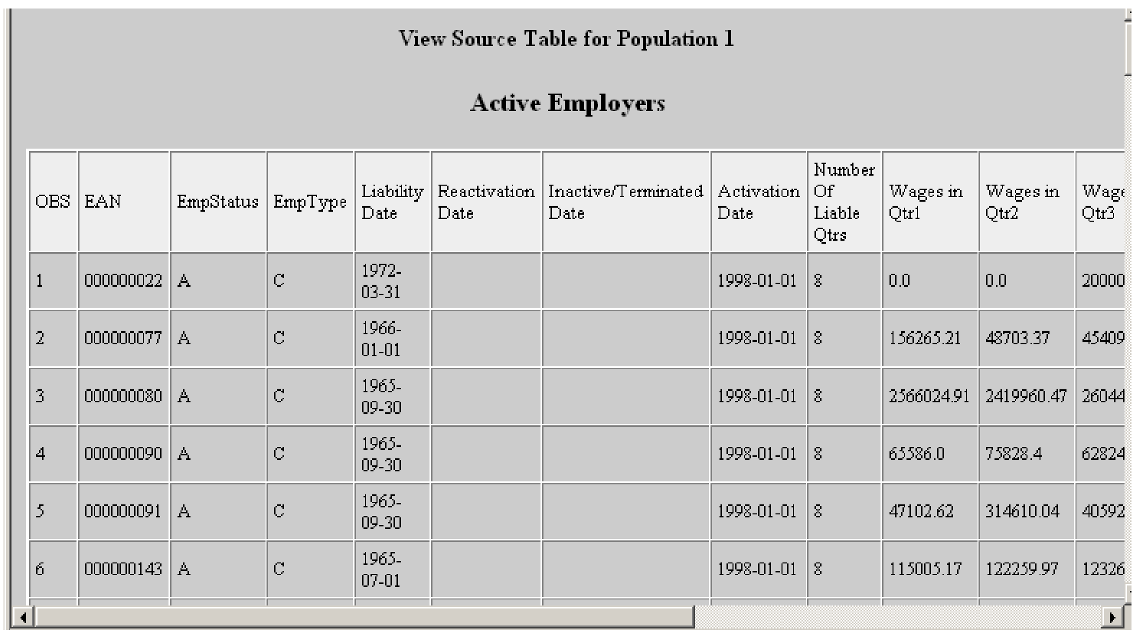
The Source Table displays all the records successfully loaded to the application for a given population, and the data elements in these records. It is accessed through the **Tax Selection Criteria** screen utilizing the *View Source Table* option.

1. From the **Tax Selection Criteria** screen select a population that has been loaded.

2. Select *View Source Table* from the *Choose A Function* drop-down box (Figure 2.2 **Tax Selection Criteria** screen).

3. Click on *Go* to view the data**.**

Figure 3.5a - View Source Table Screen (Top)



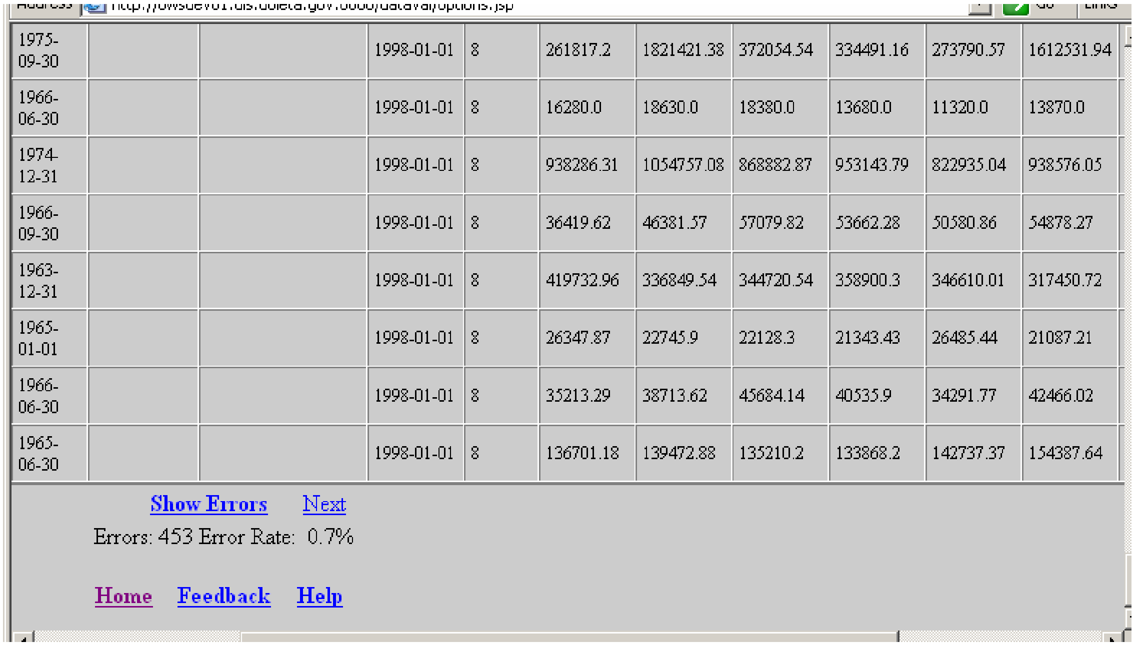
View Source Table for Population 1

Active Employers

OBS	EAN	EmpStatus	EmpType	Liability Date	Reactivation Date	Inactive/Terminated Date	Activation Date	Number Of Liable Qtrs	Wages in Qtr1	Wages in Qtr2	Wage Qtr3
1	000000022	A	C	1972-03-31			1998-01-01	8	0.0	0.0	20000
2	000000077	A	C	1966-01-01			1998-01-01	8	156265.21	48703.37	45409
3	000000080	A	C	1965-09-30			1998-01-01	8	2566024.91	2419960.47	26044
4	000000090	A	C	1965-09-30			1998-01-01	8	65586.0	75828.4	62824
5	000000091	A	C	1965-09-30			1998-01-01	8	47102.62	314610.04	40592
6	000000143	A	C	1965-07-01			1998-01-01	8	115005.17	122259.97	12326

4. The software presents data in sets of 100, so you will have to scroll through multiple screens to view populations of more than 100 records. Select the *Next* link at the bottom of the screen to view the subsequent set of records.

Figure 3.5b - View Source Table Screen (Bottom)



1975-09-30			1998-01-01	8	261817.2	1821421.38	372054.54	334491.16	273790.57	1612531.94
1966-06-30			1998-01-01	8	16280.0	18630.0	18380.0	13680.0	11320.0	13870.0
1974-12-31			1998-01-01	8	938286.31	1054757.08	868882.87	953143.79	822935.04	938576.05
1966-09-30			1998-01-01	8	36419.62	46381.57	57079.82	53662.28	50580.86	54878.27
1963-12-31			1998-01-01	8	419732.96	336849.54	344720.54	358900.3	346610.01	317450.72
1965-01-01			1998-01-01	8	26347.87	22745.9	22128.3	21343.43	26485.44	21087.21
1966-06-30			1998-01-01	8	35213.29	38713.62	45684.14	40535.9	34291.77	42466.02
1965-06-30			1998-01-01	8	136701.18	139472.88	135210.2	133868.2	142737.37	154387.64

Show Errors Next

Errors: 453 Error Rate: 0.7%

Home Feedback Help

The **View Source Table** screen indicates the number of errors detected when the extract file was loaded. In this example this population has 453 errors with an error rate of 0.7%.

- This report provides the number of records that were included in the validation count. The number of records included in the validation count plus the number of records found to be in error equals the number of records in the extract file. The *Show Errors* link on the screen allows the user to view the error records and associated error messages. If there were no errors the *Show Errors* link would be replaced with a *No Errors* message.

- The error screens can be accessed either from the *Show Errors* link at the bottom of the **View Source Table** screen or by selecting the *View Errors* function off the *Select A Function* menu.

4. Report Validation (RV)

4.1 Overview

The report validation (RV) process consists of establishing the extent to which reported numbers match summery report counts reproduced through the data validation process. This comparison process is automated and does not require additional input from state staff once an extract file has been successfully loaded to the system

Comparisons are conducted at the report item level, and pass/fail scores are calculated for individual report data elements. There is not a “total” score calculated for the complete population data set. If the error rate for one or more individual report elements is outside the tolerance level, the population-specific extract file must be corrected and re-validated in order to pass.

The computer automatically retrieves reported numbers for conducting the comparison from report files resident in a database at the state level. The population being processed defines the report(s) used for comparison, and the date information provided by the user when the extract file is imported identify the time period needed to establish which reported numbers to retrieve from the state file of 581 Reports to be used for comparison.

Summary reports produced by the software provide the results of the RV comparison. These reports can be forwarded to DOL to document the results of the validation.

While DV validation can confirm that reported counts match validation counts, DEV validation is required to verify that the extract file used for validation contains the appropriate data. Data Element Validation (DEV) performs this verification and is described in Section 5.

> NOTE: Since there are two components to validation of a given extract file (DEV and RV validation), failure of DEV validation invalidates the RV validation. Therefore, DEV validation should be passed for a given tax population before RV validation reports can be forwarded to DOL.

4.2 System Navigation

- The **View Validation Counts** screen is for informational purposes only and thus does not have a *submit* button at the bottom of the screen.

- The *Sub Pop* column on the **View Validation Counts** screen allows the user to identify the number of records assigned by the software to various sub-populations within the population data set. The sub-population identifiers in the *Sup Pop* column are highlighted on the screen and serve as links to a second screen which displays the records which have been assigned to a given sup-population.

- The **View Reported Counts** and **View Reported Dollars** Screens provide the actual comparison of reported data with validation counts. The reports provide the actual difference between the two numbers and the percent of error any observed differences represent. Report items have cumulative scores computed and are assigned the value *Pass* or *Fail* based on these results. These screens serve as summary reports providing the results of RV validations. The state can forward these reports to DOL using the Submit button on the bottom of the screen at their discretion. Submissions are transferred to DOL through overnight batch processing.

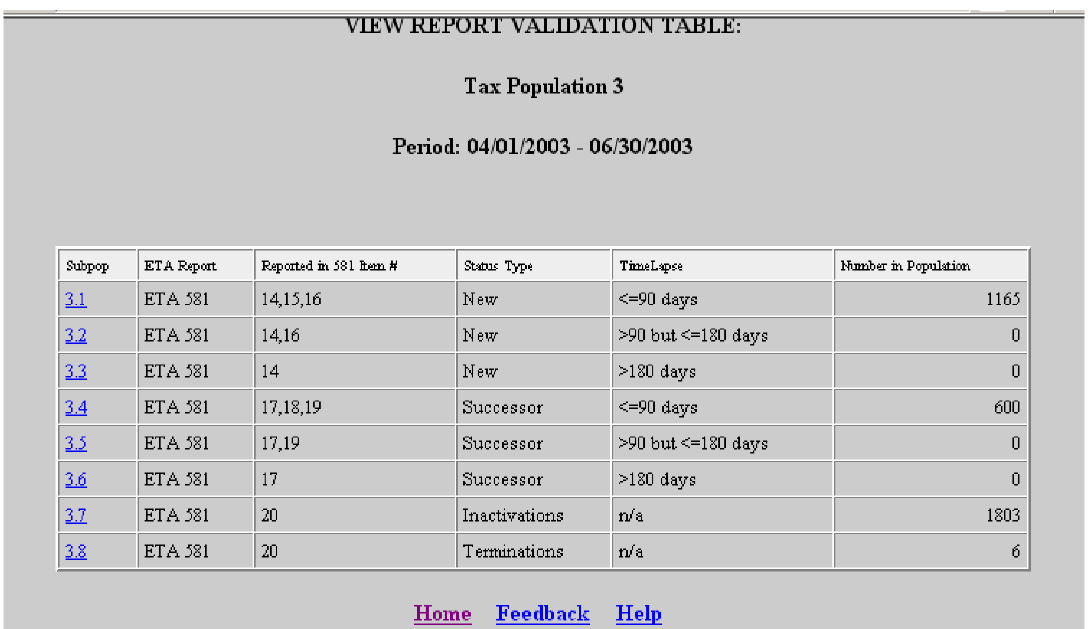
4.3 View Validation Counts

1. From the **Tax Selection Criteria** screen select a population that has been loaded.

2. Select *View Validation Counts* from the *Choose A Function* drop-down box.

3. Click on *Go.* It will bring up the **Report Validation** table:

Figure 4.1 – Report Validation Table Screen



VIEW REPORT VALIDATION TABLE:

Tax Population 3

Period: 04/01/2003 - 06/30/2003

Subpop	ETA Report	Reported in 581 Item #	Status Type	TimeLapse	Number in Population
3.1	ETA 581	14,15,16	New	<=90 days	1165
3.2	ETA 581	14,16	New	>90 but <=180 days	0
3.3	ETA 581	14	New	>180 days	0
3.4	ETA 581	17,18,19	Successor	<=90 days	600
3.5	ETA 581	17,19	Successor	>90 but <=180 days	0
3.6	ETA 581	17	Successor	>180 days	0
3.7	ETA 581	20	Inactivations	n/a	1803
3.8	ETA 581	20	Terminations	n/a	6

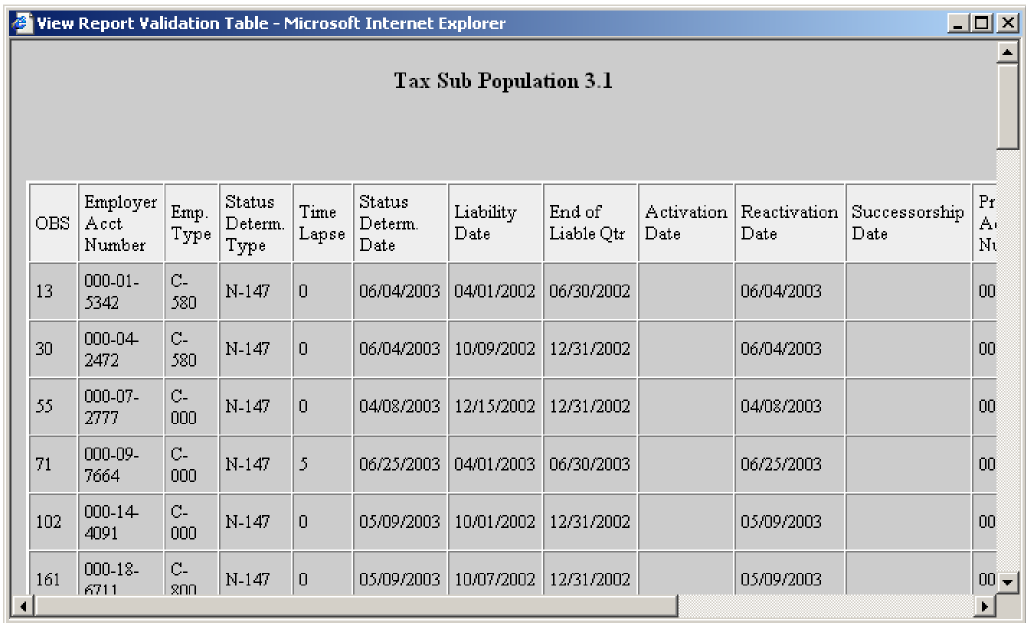
Home Feedback Help

4. This displays a summary data for each sub-population, and the total number of records and dollar amounts, where applicable, for each subpopulation. There are eight subpopulations in this population. To view the detailed records in each subpopulation, click on the *Subpop* number on the left. This will open another window displaying the detailed records for that particular sub-population (below).

4.4 View Sub-populations

This screen is accessed by clicking on the number of the sub-populations displayed on the **View Validation Counts** Screen, above. The number of the sub-population being displayed is displayed at the top of the screen.

Figure 4.2 – Sub Populations Screen



OBS	Employer Acct Number	Emp. Type	Status Determ. Type	Time Lapse	Status Determ. Date	Liability Date	End of Liable Qtr	Activation Date	Reactivation Date	Successorship Date	Pr A Nu
13	000-01-5342	C-580	N-147	0	06/04/2003	04/01/2002	06/30/2002		06/04/2003		00
30	000-04-2472	C-580	N-147	0	06/04/2003	10/09/2002	12/31/2002		06/04/2003		00
55	000-07-2777	C-000	N-147	0	04/08/2003	12/15/2002	12/31/2002		04/08/2003		00
71	000-09-7664	C-000	N-147	5	06/25/2003	04/01/2003	06/30/2003		06/25/2003		00
102	000-14-4091	C-000	N-147	0	05/09/2003	10/01/2002	12/31/2002		05/09/2003		00
161	000-18-6711	C-800	N-147	0	05/09/2003	10/07/2002	12/31/2002		05/09/2003		00

Click on the “X” in the upper right hand corner of the screen to close the screen. Click on the Home link to return to the **Tax Selection Criteria** screen.

4.5 View Reported Counts

This screen provides a summary report on the results of summary validation for a given population with reference to reported counts. Populations 1 – 3 and 3a and 5-11 include reported counts only. Populations 4 and 12 include reported dollars, as well as reported counts. Populations 13 and 14 include only reported dollars, and therefore this section does not apply to those populations.

1. From the **Tax Selection Criteria** screen select a population that has been loaded.

2. Select *Reported Counts* from the *Choose A Function* options**.**

3. Click on *Go* to view the **Report Counts** table.

4. The **Reported Counts** screen will display the report cell, description, validation counts, reported count, count difference, percent difference and whether individual report items passed or failed summary validation. (The data displayed below is test data.)

Figure 4.3a - Reported Counts Screen

Reported Counts for Population 1

Period: 01/01/2003 - 03/31/2003

Report Cell	Description	Validation Counts	Reported Counts	Difference	Percent Difference	Pass/Fail
581-101-01	Active contributory employers	65,985	393,572	327,587	496.46%	FAIL
581-101-02	Active reimbursing employers	2,188	3,299	1,111	50.78%	FAIL
Active Employers		68,173	396,871	328,698	482.15%	FAIL

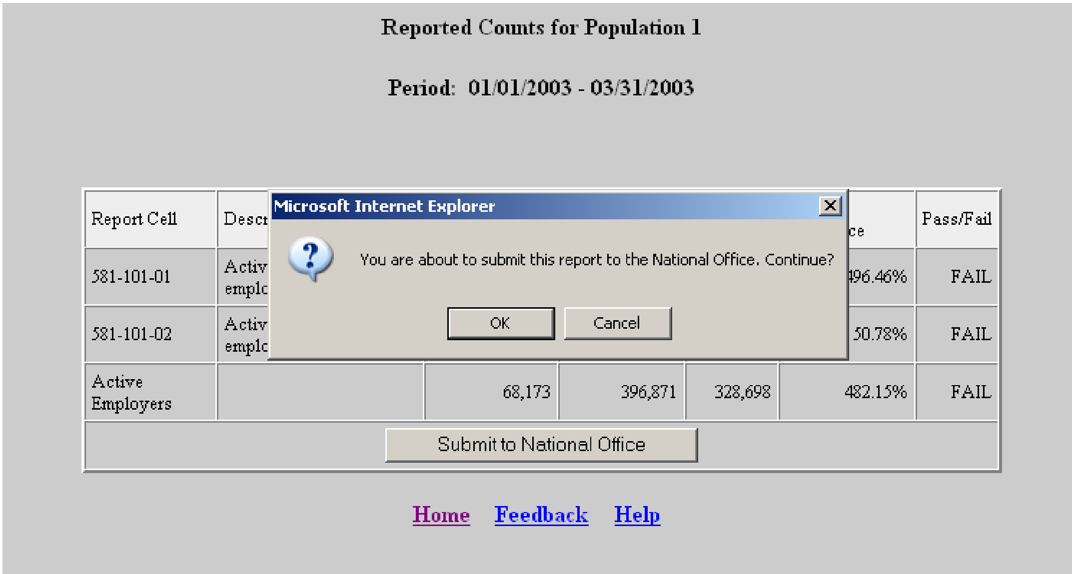
Submit to National Office

Home Feedback Help

6. After completing the Report Validation for the reported counts, click on the *Submit to the National Office* button. A warning message will appear “You are about to submit this report to the National Office. Continue?”

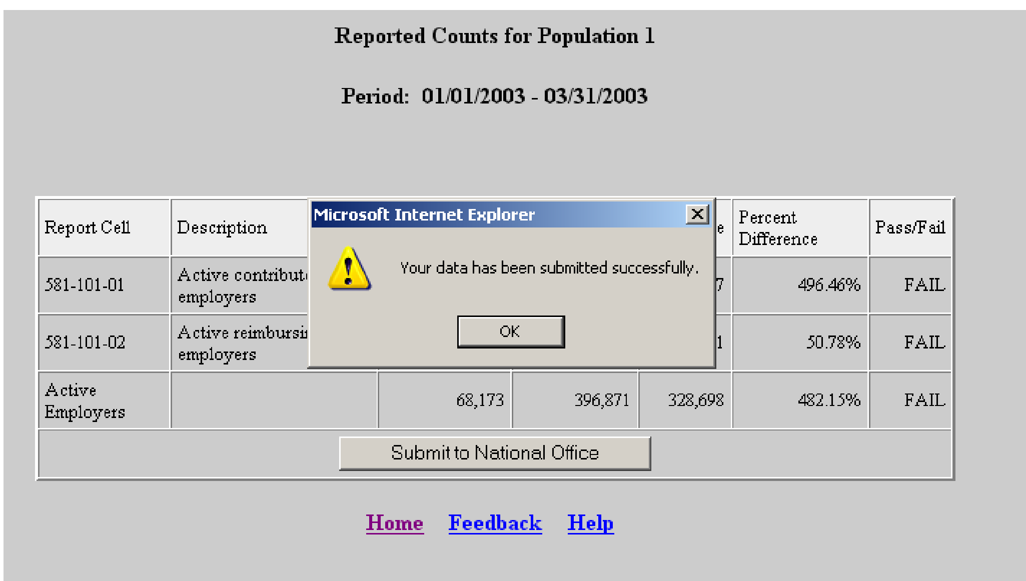
   NOTE: Summary (RV) reports should not be submitted to DOL unless a given population has first passed Data Element Validation (DEV).

Figure 4.3b - Reported Counts Screen: Submit



7. Click the *OK* button and the following screen will appear.

Figure 4.3c - Reported Counts Screen



The displayed message verifies that the report will be transferred to DOL through overnight batch processing.

4.6 View Reported Dollars

This screen provides summary results for Report Validation for populations that include reported dollar amounts. Tax populations 4 and 12 include reported dollars, Tax population 5 includes reported counts as well as reported dollars.

Reported numbers which are compared with validation counts are retrieved from state databases in the manner described for the validation of counts (above).

1. From the **Tax Selection Criteria** screen select a population that has been loaded.

2. The *Reported Dollars* option is available for tax populations 4 and 5.

3. Click on *Go* to view the **Reported Dollars** screen.

4. The **Reported Dollars** screen will display the report cell, description, validation dollars, reported dollars, difference, percent difference and individual report items passed or failed. This screen is displayed as Figures 4.4 below. (Data displayed in Figure 4.4 is test data.)

Figure 4.4 – View Reported Dollars Screen

Reported Dollars for Population 4

Period: 04/01/2001 - 06/30/2001

Report Cell	Description	Validation Dollars	Reported Dollars	Difference	Percent Difference	Pass/Fail
581-401-22	Established in RQ	153	59,996,142	59,995,989	39,213,064.71%	FAIL
581-401-23	Liquidated in RQ	316	47,968,322	47,968,006	15,179,748.73%	FAIL
581-401-24	Total Uncollectable	259	3,038,571	3,038,312	1,173,093.44%	FAIL
581-401-25	Total Removed	0	1,320,729	1,320,729	100.00%	FAIL
581-401-26	Total Balance EOQ	243	22,335,365	22,335,122	9,191,408.23%	FAIL
581-403-34	Established in RQ	0	1,220,903	1,220,903	100.00%	FAIL
581-403-35	Liquidated in RQ	131	739,544	739,413	564,437.40%	FAIL
581-403-36	Total Uncollectable	5,757	7,851	2,094	36.37%	FAIL
581-403-37	Total Removed	0	9,170	9,170	100.00%	FAIL
581-403-38	Total Balance EOQ	346	1,730,978	1,730,632	500,182.66%	FAIL

Submit to National Office

5. After reviewing Report Validation of the reported dollars, the user can choose to submit to DOL by clicking the “Submit to the National Office” button. A warning message will appear requiring the user to verify that submission is appropriate.

4.7 Comments

Comments may be attached to any data validation report being submitted to the National Office. The system provides for this by giving the users the capability to attach a separate "comments" record to reports being submitted by states. "Comments" should be forwarded by states to DOL in order to provide information state staff feels is important and relevant to the validation process.

Comments records are associated with a given validation population. These records are attached to summary and quality validation reports for a specific population when being submitted to DOL, and are logically linked with the associated reports when the data is loaded to the national data base. Comments records are created when the user selects the "View Comments" function from the main menu.

Users should understand the following points when using the Comments function:

- The size of comments records are limited to 255 characters.

- Comments associated with a given population do not distinguish between summary validation reports and quality reports. If the comments are specific to either kind of report, this should be explained in the text.

- Prior comments for a given population are still resident in the system after they have been submitted. The user should check the contents of the Comments record before submitting. If the information in the comments record is not appropriate for a given submission, the comments should be cleared or revised before submission.

- Comments should be prepared and saved prior to submission of reports. The comments will accompany the report submission.

1. From the **Tax Selection Criteria** screen select a population that has been loaded.

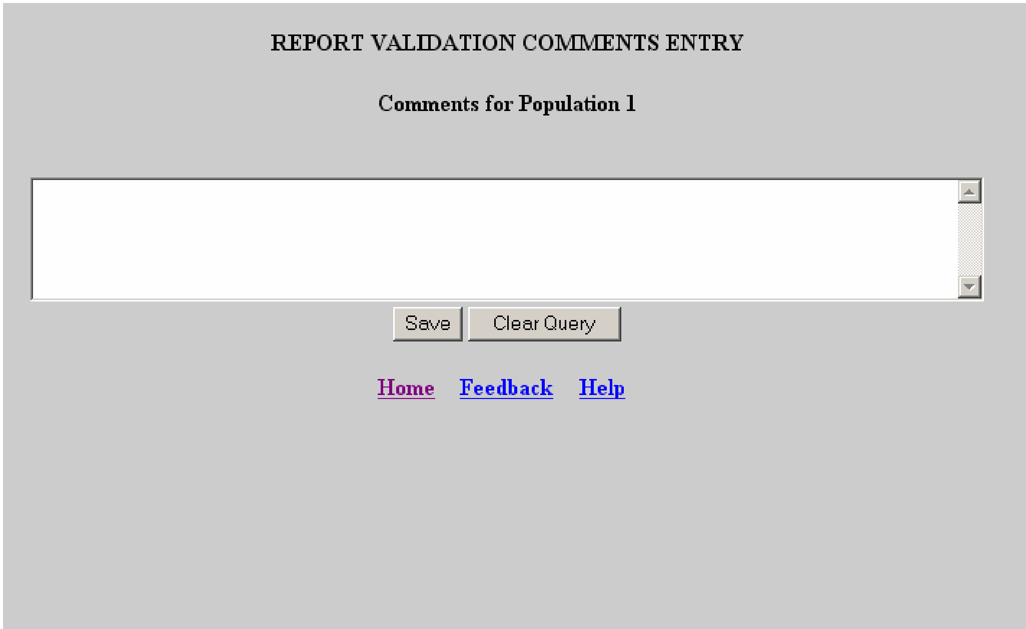
2. Select *View comments* from the *Choose A Function* options to view previous comments or enter new comments for that population.

3. Click on *Go t*o view the *Comments* table.

4. The first time you login, the comments window will be blank. You can enter your comments for that particular population.

3. Select the *Save* button to save your comments in the database. Whenever you return to this screen, the previously entered comments will be displayed. To enter a new comment, press the *Clear Query* button to remove the previous comments, and then type your new comments. Use the *Save* button at the bottom of the screen to save your new comments.

Figure 4.5 – Comments Screen



5. Data Element Validation (DEV)

5.1 Overview

There are two parts to Data Element Validation (DEV) for Tax: the investigation of samples from Tax extract files, and the analysis of tax extract files through the use of sorts. The investigation of samples serves to verify that appropriate information is included in the extract files. Sorts are used as a diagnostic tool to verify that the code values in the tax extract file are accurate and being properly counted.

The File Integrity Validation (FIV) samples drawn by the software are "minimum" samples of two records per tax sub-population. Investigators review these samples and enter the results into the database using a data-entry screen. States have passed this aspect of DEV if the sampled records are free of errors.

"Sorts" facilitate the use of range validation analysis to verify that records in the extract files have appropriate code values for certain key variables used in the Tax program. Sorts are used to parse the file into segments so that records having the same code values are grouped together. Codes which occur in the file but do not match code values expected by the software can be identified and removed or corrected. States have passed "sort" validation, or "range validation", when they have established that no more than 2% of the records in tax extract files include incorrect state-specific codes. (Sorts are not used with Tax population 5).

5.2 Wage Item Validation

This aspect of data validation consists of reviewing counts of wage record transactions which appear on the ETA 581 report to verify their accuracy. A "wage record" is the listing of an individual's earnings in covered employment. Employers are required to provide this information to the unemployment Insurance program four times per year.

Validation requires that counts from the ETA 581 be compared with reconstructed counts produced under controlled conditions. Validators test that every wage item is counted and that the count does not include corrections (counted twice), incomplete wage records or duplicate records.

5.3 System Navigation

- The **View Sample Validation** screen provides access to the sample worksheets (**Sample Validation** Screen).

- Users should be aware that loading the same population twice should not affect the results of RV validations. However, DEV validation includes samples, and sampling procedures tend to produce different sets of records each time a sample is drawn. Thus it is important that the user "save" both the original sample and the results of sample investigations to prevent the loss of work if the user accidentally exits the system or loses sample data for any other reason (i.e., a system crash).

- Several buttons are provided on the **Sample Validation** screen. These are explained below:

  Save Button. This saves information displayed on the screen within the application. Samples and the complete or partial results of sample investigations should be saved in tables set up within the DV software.

  Save As Button. This allows the user to save data sets at any location accessible from the workstation. Summary results from the validation of populations and the extract file itself should be should be saved to satisfy audit requirements.

  Save and Submit button. This saves the data to tables, which are part of the application, and submits the report to DOL using conventions set up for transfer of reports through the Sun interface in overnight batch processing.

  Print Preview button. This allows the user to view an image of the document as it will be printed before submitting a print request. This is useful because the user may need to make adjustments before printing the sample investigation worksheets due to the format of the document.

  Print button. This sends an image of the displayed screen to a local printer. The *print* option allows the user to specify the number of records to be printed by providing *start row* and *end row*.

  Print button. This routs an image of the displayed screen to a local printer.

5.4 View Samples

This screen identifies samples which have been drawn from a given tax population. It also provides access to the **Sample Validation** screen which provides data entry capability to store the results of sample investigations.

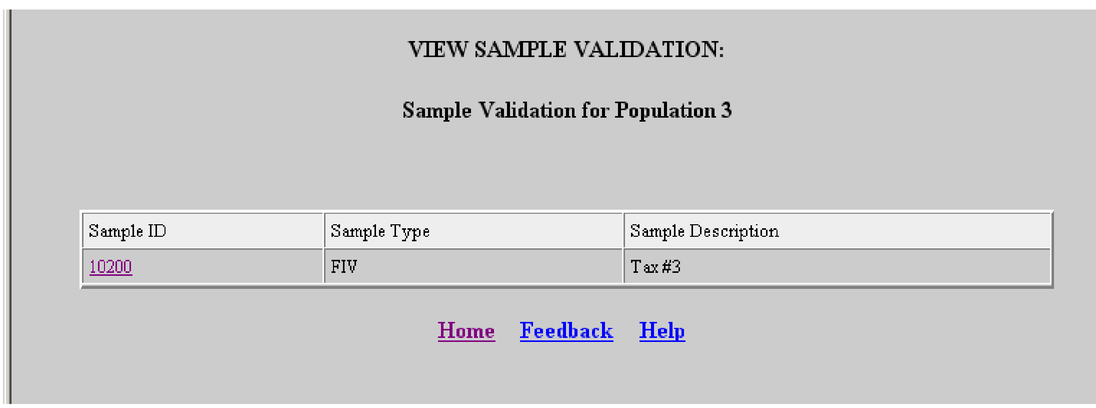
1. From the **Tax Selection Criteria** screen, select a population that has been loaded.

2. Select *View Samples* from the *Choose A Function* options.

3. Click on *Go* to view the **View Sample Validation** table. It will open a screen showing the samples that are selected for the population. Samples are referenced to specific sub-populations.

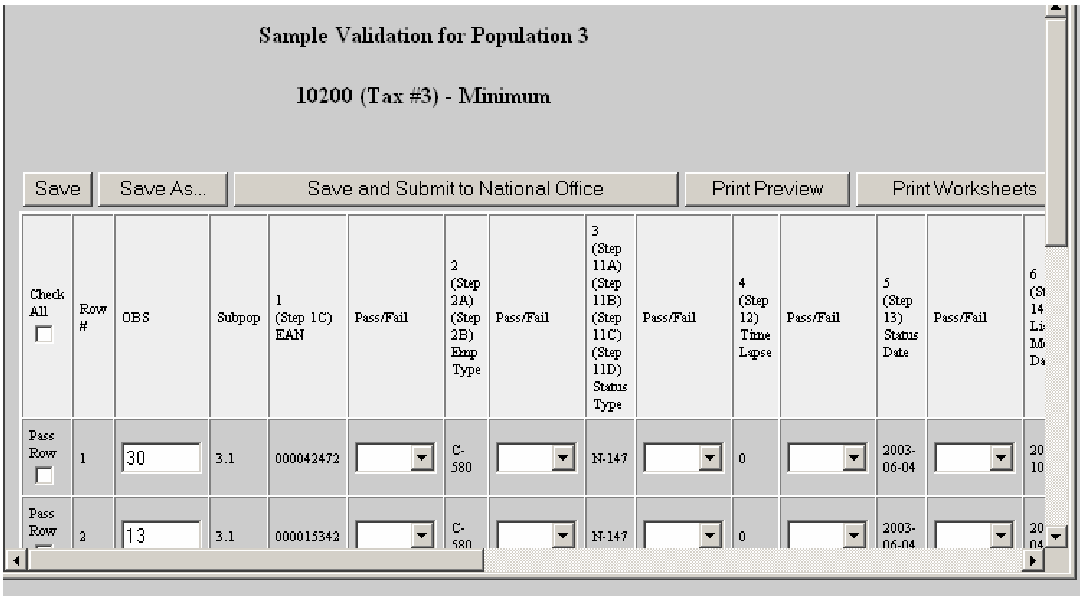
NOTE: The *Sample ID* number identifies the specific sample. *Sample Type* provides a description of the type of sample. All tax populations have one sample each and they are all FIV (File Integrity Validation Samples: minimum samples of two per sub-population).

Figure 5.1 – View Sample Validation Screen



4. The highlighted numbers in the *Sample ID* column serve as links to the **Sample Validation** Screen. Clicking on the highlighted numbers in the *Sample ID* column will open a new screen displaying records which have been selected for that sample: **Sample Validation** Screen (below).

Figure 5.2 – Sample Validation Screen



NOTE: The **Sample Validation** Screen displays records selected by the sample routine which are to be investigated. The results of the investigation are data-entered into the system using this screen. Summary results of this investigation are displayed at the bottom of the screen. Summary results for the sample are transmitted to DOL when the user selects the *Save and Submit to National Office* option. Detail information from individual records is not transmitted to DOL. This means that sensitive information, such as SSNs, stays at the state level.

5. Click the *Print Worksheet***s** button to print the sample worksheets for the set of individual records which constitute the sample.

NOTE: Users are urged to print this screen as an aid in the investigation. Information gathered through the investigation and noted on the printed forms can be data-entered at one time, avoiding any time-out problems with the application

6. This screen facilitates sample investigations by providing a box next to each data element in the sampled records for data entry, accompanied by a drop-down menu allowing the user to select “pass” or “fail” for each element. Each element is checked against the state database. The step numbers on the screen refer to the step in the validation handbook. Follow the rules for each step, as specified in Module 3 in the Data Validation Handbook, by checking all necessary screens and other documentation to determine whether each element is valid. Select “Pass” or “Fail” depending on whether the element in the record corresponds to the data in the database.

NOTE: It is important to save both the sample and the results of investigations during the process to assure that data is not lost. Data can be saved in internal tables using the *Save* button.

7. The user has the option to select *Pass Row* or *Check All*, to set values for individual records, or for the entire sample. Use of this option should be approached with caution since elements set to “fail” will default to “pass” if this option is used.

8. Sample investigations are complete when the records in the sample have results in all the data entry boxes. Completed sample investigations can be transmitted to DOL through use of the *Save and Submit* button.

NOTE: Sample validations cannot be submitted to DOL unless all the data-input boxes for a given sample are completed. Incomplete investigation results cannot be submitted.

5.5 View Data Element Sorts

Methodology

Range validation (sorts) is a series of tests to determine whether the generic primary codes used to assign transactions (e.g., Contributory or Reimbursing Employers in Populations 1 and 2) are accurately supported by state-specific secondary codes or specific ranges of Employer Account Numbers (EANs). If a state's database does not have more than one state-specific code for a given generic code, or does not use EAN ranges, the following tests do not apply. In the software they are called Data Element Validation (DEV)/Sorts. A data element passes DEV if no more than 2% of the sorted transactions include an incorrect state-specific code. An extract file (a) must pass all applicable DEV tests and (b) have no FIV samples with errors before the RV can pass. Range tests are explained in Module 2.3 of the Tax Handbook, pages 2-14 through 2-19. These tests do not apply to Population 5.

The software provides a data entry screen on which the validator records for each sort: the number of employers or transactions subjected to the range test and the number of transactions that are "out of range." When *Save* or *Submit to National Office* is clicked, the software computes the percentage that is out of range. Completing this function thus requires the validator to know (1) the number of employers or transactions subjected to the range test; and (b) the number out of range.

The total number of employers is obtained from the **View Validation Counts** screen by summing the number in the specified set of subpopulations. The Table on the following page illustrates the relationship among Tax populations, sorts, and sub-populations. *Test Data Element* identifies the data element used by the software to perform the sort.

Table 5.1: Tax Populations, Sorts and Sub-populations

Population	*Sort*	*Subpopulations Examined*	*Test Data Element*
1	S1.1	1.1 (Contributory Employers)	EAN
1	S1.2	1.2 (Reimbursing Employers)	EAN
1	S1.3	1.1 + 1.2 (All employers)	Employer Status Indicator
1	S1.4	1.1 (Contributory Employers)	Employer Type Indicator
1	S1.5	1.2 (Reimbursing Employers)	Employer Type Indicator
2	S2.1	2.1 – 2.8 (Contributory Employers)	EAN
2	S2.2	2.9 – 2.16 (Reimbursing Employers)	EAN
2	S2.3	2.1 – 2.8 (Contributory Employers)	Employer Type Indicator
2	S2.4	2.9 – 2.16 (Reimbursing Employers)	Employer Type Indicator
3	S3.1	3.1 – 3.3 (New Status Det.)	Status Determination Type
3	S3.2	3.4 – 3.6 (Successor Status Det.)	Status Determination Type
3	S3.3	3.7 (Inactivation Det.)	Status Determination Type
3	S3.4	3.8 (Termination Det.)	Status Determination Type
4	S4.1	4.1, 4.9 (Establishment Transaction)	Transaction Type Indicator
4	S4.2	4.2, 4.10 (Liquidation Transaction)	Transaction Type Indicator
4	S4.3	4.3, 4.4, 4.11, 4.12 (Uncollectible Transactions)	Transaction Type Indicator
Source: Handbook, Exhibit 2.8.			

There are two types of Range Tests/Data Element Sorts to determine which employers or transactions are out of range. The software performs them differently.

❖ EAN Range Tests. If your state assigns a certain range of EAN numbers to Contributory Employers and another to Reimbursing numbers, these tests apply.

  - They are S1.1 and S1.2 for Population 1, S2.1 and S2.2 for Population 2.

  - How the software works: When you select one of the EAN tests the software prepares to do a query against only those transactions in the appropriate set of subpopulations. E.g., selecting S1.1 prepares a query against subpopulation 1.1 (Contributory Employers); S1.2 prepares a query against subpopulation 1.2, Reimbursing employers.

  - The software give a choice of three possible queries:
    - *Starts With* retrieves all employers whose EANs begin with the sequence you identify, e.g., 0079. Enter 0079 in the first box.
    - *Ends With* retrieves all employers whose EANs end with the sequence you identify, e.g., 9999. Enter 9999 in the second box.
    - *Is Between* allows you to specify a range starting with an EAN number in the first box and an ending value in the second box.

        - This function can also be used in lieu of *Starts With*, e.g., to identify the series of 9-digit EANs beginning with 0079, because they would end with the value 007999999.

  - The query returns the set of employers that falls within the specified range.
    - The number of errors is the total number of records in the set against which the query is run less the number in range.
    - If the number returned by the query exceeds 10,000—very likely in the case of S1.1 and S2.1--Contributory Employers—the software displays the first 10,000, beginning with the lowest EAN.
      - To determine the total number of records in such a case requires repeated queries to using the *Is Between* function, starting with EAN+1 from the preceding query.
    - A more efficient approach in the case of S1.1 and S2.1 is to do queries for records at the upper and lower limit of the EAN range.
      - E.g., if the EAN range is 007900000 to 007999999, run a query for records beginning with 000000000 to 007899999, and another for 008000000 and above. Any out-of-range records returned are errors.

- <u>Distributions/Cross-tabulations</u>. The other sorts—S1.3-S1.5, S2.3-S2.4, and Population 3 and 4 sorts—involve instances where a state may have multiple codes for active employers, employer type, or types of transactions. In these situations, the software presents a combined distribution or cross-tabulation of contributory and reimbursing employers, or all status determinations, etc., by the state's supplementary codes.
  - An example is the following screen shot for sorts S2.3 and S2.4 of Population 2 (employer type).

  NOTE: Because this combines all the sorted kinds of transactions, clicking on S2.3 and S2.4 brings up the same screen.

  - The total number of transactions or employers sorted is easily obtained by summing all the counts for Contributory (65,954) and Reimbursing (2,177) from the screen and comparing those with validation counts.
  - The number of errors can be seen at a glance as the number of employers with inappropriate codes. If, in the screen example below, code 750 were a code for reimbursing employers, S2.4 would have 6,007 errors out of 65,954. If all codes from 000 through 750 are appropriate for Contributory employers and 999 for Reimbursing, however, this population shows no errors for S2.3 and S2.4.
  - Click on the left-hand link for a distribution to bring up the detail records for examination. As with the EAN test, if the number of records in a category exceeds 10,000, the software retrieves the records with the lowest 10,000 EANs first.

- The software provides no means for retrieving more than the first 10,000 records in a set. In the example below, one can see only part of the detail records for Contributory-000 and Contributory-100.

Data Element Distribution for Tax population 2

Employer Type	Count
Contributory-000	28886
Contributory-010	2441
Contributory-020	926
Contributory-030	1174
Contributory-040	724
Contributory-050	1443
Contributory-060	582
Contributory-080	1119
Contributory-100	15171
Contributory-110	632
Contributory-140	1058
Contributory-170	948
Contributory-230	1564
Contributory-310	1191
Contributory-430	2088
Contributory-750	6007
Reimbursing-999	2177

Home **Feedback** **Help**

NOTE: States that are not required to perform sorts as part of Tax Data Validation should explain this to the National Office using the Tax Comments function. States are not required to perform analysis of tax extract files using sorts if: a) they do not have more than one state-specific code for a given generic code, or b) do not identify employer *type* using EAN ranges.

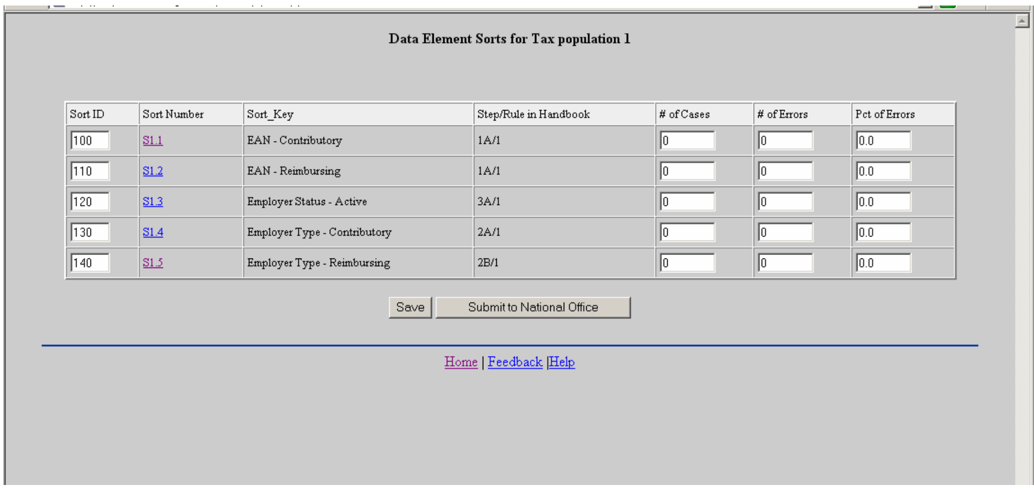
Procedure

1. From the **Tax Selection Criteria** Menu screen select a population that has been loaded.

2. Select **View Data Element Sorts** from the drop down box *Choose a Function.*

3. Click on *Go* to view the **View Data Element Sorts** table; it brings up the following screen:

Figure 5.3 – Data Element Sorts Screen



Data Element Sorts for Tax population 1

Sort ID	Sort Number	Sort_Key	Step/Rule in Handbook	# of Cases	# of Errors	Pct of Errors
100	S1.1	EAN - Contributory	1A/1	0	0	0.0
110	S1.2	EAN - Reimbursing	1A/1	0	0	0.0
120	S1.3	Employer Status - Active	3A/1	0	0	0.0
130	S1.4	Employer Type - Contributory	2A/1	0	0	0.0
140	S1.5	Employer Type - Reimbursing	2B/1	0	0	0.0

Save | Submit to National Office

Home | Feedback |Help

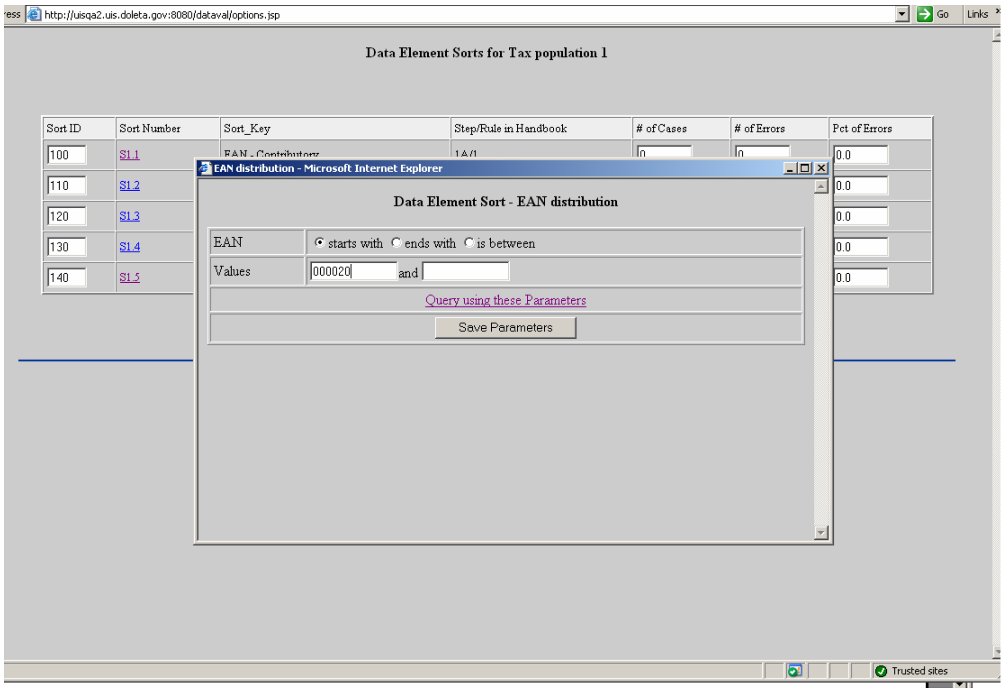
The Sort options for Populations 1 and 2 provide two different kinds of sorts.

- The EAN sort, if applicable, uses ranges of the *EAN* to indicate whether the employer is contributory or reimbursing. For Population 1, these are S1.1 (contributory employers) and S1.2 (reimbursing employers.)

- The other sorts - for Population 1, sorts S1.3, S1.4, S1.5 - use supplementary codes: Sort S1.3 uses codes on the *Active field* to confirm whether the employer is active; S1.4 and S1.5 uses codes on the *Employer Type* field to confirm whether the employer is contributory or reimbursing.

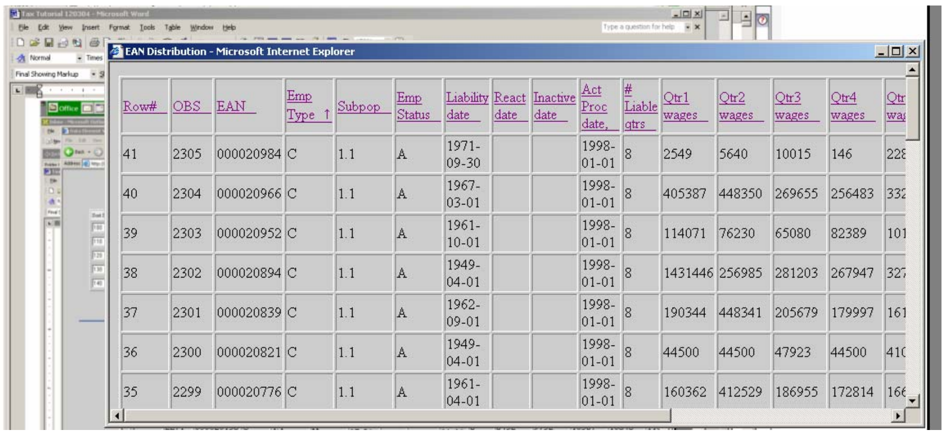
The EAN sort allows the user to specify a range of values at the beginning, end, or within the *EAN* sequence that the state uses to indicate whether an employer is contributory or reimbursing. Selecting *S1.1,* assuming that the code for contributory begins with 000020, displays the following screen:

Figure 5.4 – Data Element Sorts – EAN Distribution



4. The *Query using these Parameters* button produces results displayed on the following screen. You can click on the column headings - in this case *Emp Type* - to sort the file to identify any non-contributory employers based on the *Employer Type* variable.

Figure 5.5 – Data Element Sorts Screen

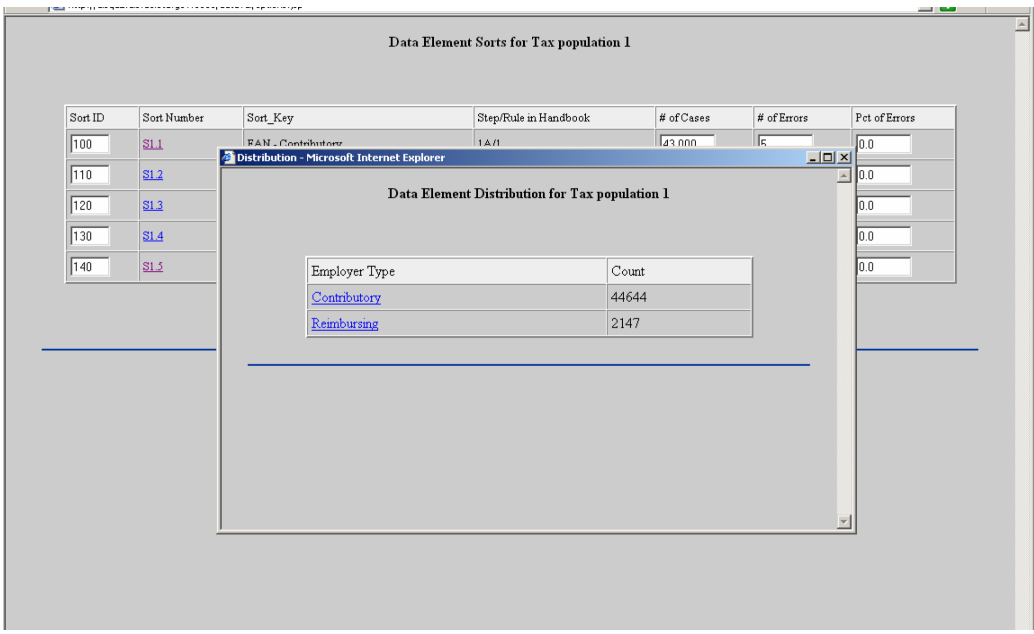


Row#	OBS	EAN	Emp Type	Subpop	Emp Status	Liability date	React date	Inactive date	Act Proc date	# Liable qtrs	Qtr1 wages	Qtr2 wages	Qtr3 wages	Qtr4 wages	Qtr wa
41	2305	000020984	C	1.1	A	1971-09-30			1998-01-01	8	2549	5640	10015	146	228
40	2304	000020966	C	1.1	A	1967-03-01			1998-01-01	8	405387	448350	269655	256483	332
39	2303	000020952	C	1.1	A	1961-10-01			1998-01-01	8	114071	76230	65080	82389	101
38	2302	000020894	C	1.1	A	1949-04-01			1998-01-01	8	1431446	256985	281203	267947	327
37	2301	000020839	C	1.1	A	1962-09-01			1998-01-01	8	190344	448341	205679	179997	161
36	2300	000020821	C	1.1	A	1949-04-01			1998-01-01	8	44500	44500	47923	44500	410
35	2299	000020776	C	1.1	A	1961-04-01			1998-01-01	8	160362	412529	186955	172814	166

5. Take the total number of contributory employers from the **Validation Count** screen for subpopulation 1.1 and enter that in the *# of Cases* field on the **Data Element Sorts** screen. Enter the number of records in error in the *# of Errors* field after examining the sorted records.

6. The software works the same for other non-EAN sorts (e.g., Sorts S1.3, S1.4, S1.5). It produces a distribution of the total of contributory and reimbursing employers by supplementary codes. Clicking on *S1.4* produces the following screen:

Figure 5.6 – Data Element Distribution for Tax

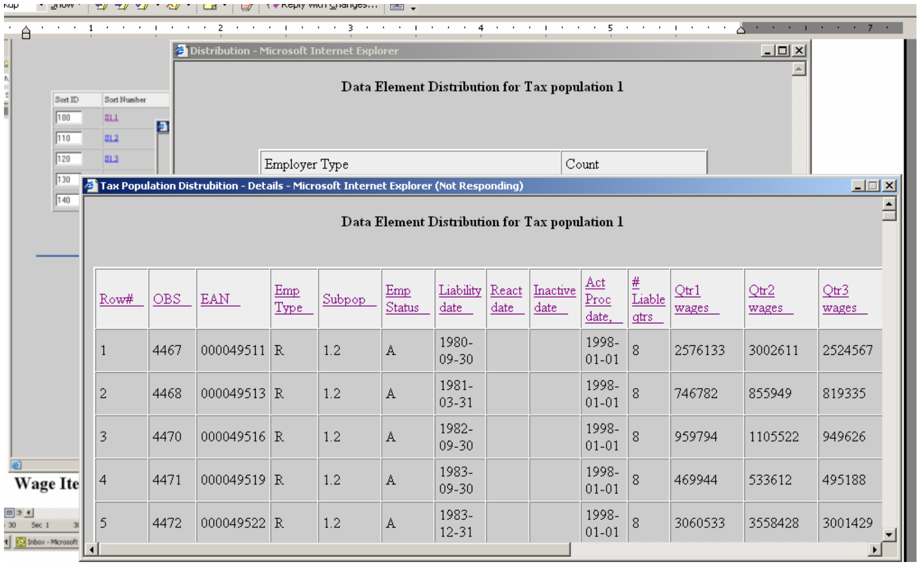


7. Clicking on *Reimbursing* brings up the following **Data Element Distribution** screen. This screen facilitates analysis of the extract file by providing the capacity to sort the file with reference to key variables. Sub-populations within the extract file can be examined for errors by sorting on the variable of interest – *Emp Type* in Figure 5.7.

- Because of the limitations of the underlying environment, if the number of records exceeds 10,000, the software will bring up only 10,000 records.

- The sort function works very slowly for large data sets.

Figure 5.7 – Data Element Distribution of Tax



8. Enter the number of records sorted in the *# of cases* field of the **Data Element Sort** screen (Figure 5.3), and the number of records in error in the *# of Errors* field. The Save button saves these results in the internal database.

9. Data are transmitted to the National Office using the *Submit to National Office* button.

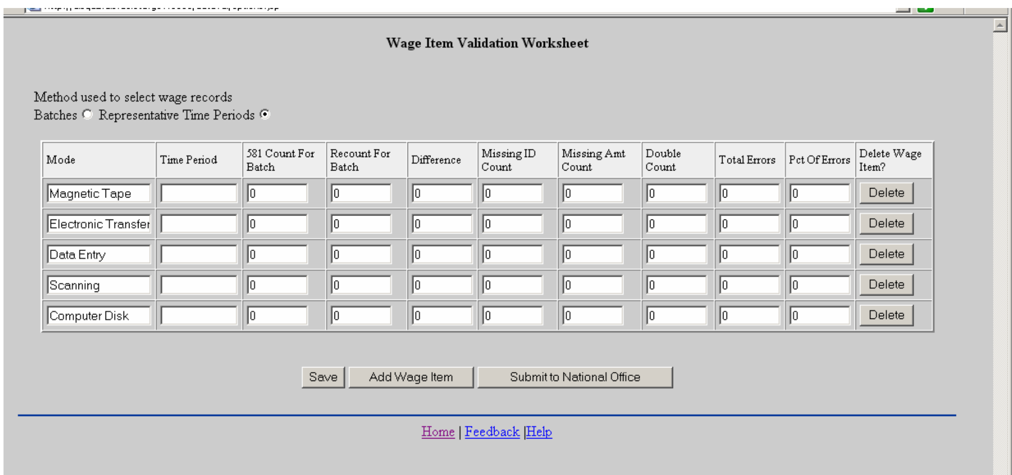
5.6 View Wage Item Validation

Wage Item Validation requires a review of wage information received from employers to verify that ETA 581 counts of this activity are accurate. A detailed explanation of the validation of wage items is provided in Module 5 of the Unemployment Insurance Data Validation Handbook for tax validation.

1. From the **Tax Selection Criteria** Menu screen select *View Wage Item Validation* from the drop-down box *Choose A Function.*

2. Click on *Go* to view the **Wage Item Validation** table. The following screen is displayed:

Figure 5.8 – Wage Item Validation Screen



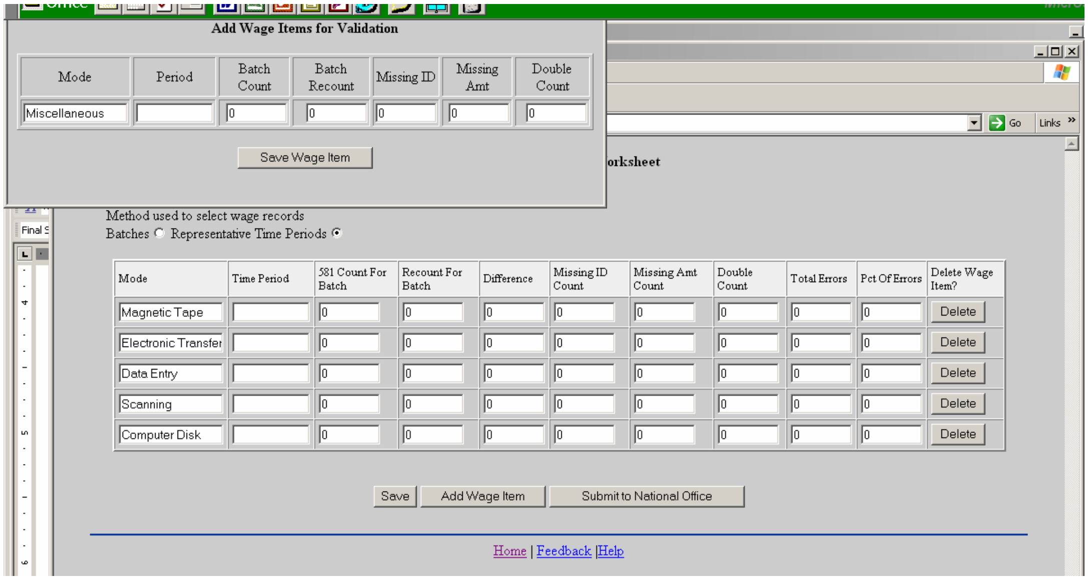
3. Indicate the way the items are selected, either Batch or Representative Time Period.

4. Enter the date used. The *Time Period* field will accept any date, but the format must be: DD/MM/YYYY.

5. The software provides for five standard methods for the receipt of wage items. The *Delete Wage Item* button allows items to be deleted to reflect state procedures. Additional items can be added using the *Add Wage Item* button. Enter the counts for the applicable time period that are reflected in a given ETA 581 Report in the *581 Count* field. Enter the recount for that category in the *Recount* field. The *Difference* and *Pct of Errors* is automatically generated when the user presses the *Save* or *Submit* buttons. Information entered in the *Missing ID*, *Missing Amt* and *Double Count* fields is for diagnostic purposes only and does not contribute to the percent of error calculations.

6. Add new rows, if required, by using the *Add Wage Items* button, e.g., if your state receives wage items in other modes, or to reverse an inadvertent deletion. It brings up the following screen:

Figure 5.9 – Wage Item Validation Screen



7. Clicking *Save Wage Item* will add the row to the **Wage Item** table. You must give the row a name, e.g., *Miscellaneous*.

8. Clicking on the Submit to National Office transfers this data to DOL through overnight batch processing.

5.7 Comments

Users my append comments to RV and DEV Tax reports being submitted to DOL. Comments are population-specific. When the *Submit* option is used, the version of the comments file resident in the software for the population being submitted will be appended to the report and transmitted to DOL. Information provided about the Comments function in Section 4, page 26, applies here also.